

Conference Abstract

Integrating ABCD and DarwinCore: Toward a better foundation for biodiversity information standards

Stanley Blum[‡], Katharine Barker[§], Steven J Baskauf[‡], Walter G. Berendsohn[¶], Pier Luigi Buttigieg[#], Markus Döring[□], Gabriele Droege[¶], David Fichtmueller[¶], Falko Glöckler[«], Anton Güntsch[¶], Robert Guralnick[»],[^],[^], Jana Hoffmann[«], Niels Klazenga[‡], James Macklin[?], Paul J. Morris[¶], Deborah L Paul[¢], Mareike Petersen[«], Tim Robertson[□], Joel Sachs[?], David Peter Shorthouse[?], Ramona Walls^ℓ, John Wieczorek[‡], Paula F F Zermoglio^P

[‡] Biodiversity Information Standards (TDWG), San Francisco, United States of America

[§] National Museum of Natural History, Smithsonian Institution, Washington, D.C., United States of America

[‡] Vanderbilt University Libraries, Nashville, Tennessee, United States of America

[¶] Freie Universität Berlin, Botanic Garden and Botanical Museum, Berlin, Germany

[#] Alfred-Wegener-Institut für Polar und Meeresforschung, Bremen, Germany

[□] Global Biodiversity Information Facility, Copenhagen, Denmark

[«] Museum für Naturkunde Berlin, Leibniz Institute for Evolution and Biodiversity Science, Berlin, Germany

[»] Vertnet, Florida, United States of America

[^] University of Colorado, Boulder; University of Colorado Museum of Natural History, Boulder, United States of America

[^] Univ. of Florida, Gainesville, United States of America

[‡] Royal Botanic Gardens Victoria, Melbourne, Australia

[?] Agriculture and Agri-Food Canada, Ottawa, Canada

[¶] Museum of Comparative Zoology, Harvard University, Cambridge, MA, United States of America

[¢] Florida State University, Tallahassee, United States of America

^ℓ Bio5 Institute, University of Arizona, Tucson, United States of America

[‡] Museum of Vertebrate Zoology, University of California, Berkeley, United States of America

^P Instituto de Ecología, Genética y Evolución de Buenos Aires (IEGEBA-CONICET), University of Buenos Aires, Buenos Aires, Argentina

Corresponding author: Stanley Blum (stanblum@gmail.com)

Received: 19 Jun 2019 | Published: 02 Jul 2019

Citation: Blum S, Barker K, Baskauf S, Berendsohn W, Buttigieg P, Döring M, Droege G, Fichtmueller D, Glöckler F, Güntsch A, Guralnick R, Hoffmann J, Klazenga N, Macklin J, Morris P, Paul D, Petersen M, Robertson T, Sachs J, Shorthouse D, Walls R, Wieczorek J, Zermoglio P (2019) Integrating ABCD and DarwinCore: Toward a better foundation for biodiversity information standards. Biodiversity Information Science and Standards 3: e37491. <https://doi.org/10.3897/biss.3.37491>

Abstract

For the last 15 years, Biodiversity Information Standards ([TDWG](https://www.tdwg.org/)) has recognized two competing standards for organism occurrence data, [ABCD](https://www.tdwg.org/abcd/) (Access to Biological

Collections Data; Holetschek et al. 2012) and [DarwinCore](#) (Wieczorek et al. 2012). These two representations emerged from contrasting strategies for mobilizing information about organism occurrences (also commonly called species occurrence data). ABCD was capable of representing details of more kinds of information, but was necessarily more complicated. DarwinCore, on the other hand, was simpler but more limited in its ability to represent data of different kinds and formats. TDWG endorsed both standards because the different projects and communities that generated them remained dedicated to their different strategies and tool sets, and the Global Biodiversity Information Facility ([GBIF](#)) developed the ability to integrate data published in either standard.

Since their inceptions, DarwinCore and ABCD have become more similar. DarwinCore has gotten more complicated through the addition of terms and has begun to assign terms to classes. ABCD is now expressed in [RDF](#) (Resource Description Framework), potentially enabling re-use of terms with alternative structures among classes. At the same time, methodologies for conceptual modeling and representing complex scientific data have continued to evolve. In particular, a suite of modeling and data representation methods related to [linked data](#) and the [semantic web](#), i.e., RDF, [SKOS](#) (Simple Knowledge Organization System), and [OWL](#) (web Ontology Language), promise to make it easier for us to reconcile shared concepts among different representations or schemas.

A mapping between ABCD 2.1 and DarwinCore has existed since before 2005.*1 ABCD 3.0 and DarwinCore are both now represented in RDF. In addition, the [BioCollections Ontology](#) (BCO) covers many of the shared concepts and is derived from the [Basic Formal Ontology](#) (BFO), an upper level ontology that has oriented many other biomedical ontologies. Reconciling ABCD and DarwinCore through alignment with BCO (in the [OBO Foundry](#); Smith et al. 2007) would better connect TDWG standards to other domains in biology. We appreciate that many working scientists and data managers perceive ontologies as overly complicated. To mitigate the steep learning curve associated with ontologies, we expect to create simpler application profiles or schemas to guide and serve narrower communities of practice within the wider biodiversity domain. We also plan to integrate the current work of the Taxonomic Names and Concepts Interest Group and thereby eliminate the redundancy between DarwinCore and [Taxonomic Concepts Transfer Schema](#) (TCS; Kennedy et al. 2006).

At the time of this writing, we have only agreements from the authors (i.e., conveners of relevant TDWG Interest Groups and other key stakeholders) to collaborate in pursuit of these common goals. In this presentation we will give a more detailed description of our objectives and products, the methods we are using to achieve them, and our progress to date.

Keywords

metadata, data model, schema integration, ontology, RDF, OWL, natural history collections, taxonomic concepts, semantic web, linked data

Presenting author

Stanley Blum

Presented at

Biodiversity_Next 2019

References

- Holetschek J, Dröge G, Güntsch A, Berendsohn WG (2012) The ABCD of primary biodiversity data access. Plant Biosystems 146 (4): 771-779. <https://doi.org/10.1080/11263504.2012.740085>
- Kennedy J, Hyam R, Kukla R, T. Paterson (2006) Standard Data Model Representation for Taxonomic Information. OMICS: A Journal of Integrative Biology. Volume 10 (2). URL: <http://doi.org/10.1089/omi.2006.10.220>
- Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, Goldberg LJ, Eilbeck K, Ireland A, Mungall CJ, Leontis N, Rocca-Serra P, Rutterberg A, Sansone S, Scheuermann RH, Shah N, Whetzel PL, Lewis S (2007) The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nature Biotechnology 25 (11): 1251-1255. <https://doi.org/10.1038/nbt1346>
- Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, Giovanni R (2012) Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. PLoS. ONE 7 (1): 29715. URL: <https://doi.org/10.1371/journal.pone.0029715>

Endnotes

- *1 See the BioCASE provider software version history. The ABCD to Darwin Core mapping is referenced in version 2.1.0, which preceded 2.2.0 and was issued 2005-07-18. <https://wiki.bgbm.org/bps/index.php/VersionHistory>. Retrieved 2019-05-09.